```
/*
*******************************************************************************
PROJECT: Global Breast Cancer Burden 2022
AUTHOR: Andy Harper
PI: Miranda Fidler-Benaoudia
START DATE: February 27, 2024
LAST UPDATED: November 11, 2024
*******************************************************************************
*/
clear all
* Define the file locations before doing anything...
global main "R:/Research/SSU AH/2024 Projects/MFB/Atlas Breast Cancer Chapter"
global data "${main}/Finalized Data"
global figures "${main}/Finalized Figures"

**# EAPC DATA GENERATION
/*
*******************************************************************************
STEP 1
Generating the EAPC dataset...
*******************************************************************************
*/
{
* Clean HDI dataset before doing anything else...
clear all
import delimited "${data}/UNDP Full Data.csv", case(lower)
keep country hdi_2013
rename country label_country
replace label_country = "Korea, Republic of" if label_country == "Korea (Republic of)"
replace label_country = "Turkey" if label_country == "Türkiye"
replace label_country = "USA" if label_country == "United States"
* Need subregional data (Puerto Rico = USA, France (Martinique) = France, UK for all
subregions)
insobs 6
replace label_country = "Puerto Rico" if _n == 207
replace label_country = "France, Martinique" if _n == 208
replace label_country = "UK, England" if _n == 209
replace label_country = "UK, Scotland" if _n == 210
replace label_country = "UK, Northern Ireland" if _n == 211
replace label_country = "UK, Wales" if _n == 212
replace hdi_2013 = 0.922 if label_country == "Puerto Rico"
replace hdi_2013 = 0.888 if label_country == "France, Martinique"
replace hdi_2013 = 0.923 if inlist(label_country, "UK, England", "UK, Scotland", "UK,
Northern Ireland", "UK, Wales")
save "${data}/Manuscript HDI Data.dta", replace

* Clean ASR dataset...
clear all
import excel "${data}/Manuscript ASR Data.xlsx", case(lower) first
* Fix country labels first...
replace label_country = "France" if label_country == "France (metropolitan)"
replace label_country = "Turkey" if label_country == "TÃ¼rkiye"
replace label_country = "Netherlands" if label_country == "The Netherlands"
labmask code_age, val(label_age)
labmask code_type, val(label_type)
labmask code_country, val(label_country)
labmask code_sex, val(label_sex)
* Keep 20-year range...
keep if inrange(year, 1998, 2017)
* Save dataset...
save "${data}/Manuscript ASR Data (1998-2017).dta", replace

* Generate EAPC dataset...
clear all
import excel "${data}/Manuscript ASR Data.xlsx", case(lower) first
* Fix country labels first...
replace label_country = "France" if label_country == "France (metropolitan)"
replace label_country = "Turkey" if label_country == "TÃ¼rkiye"
replace label_country = "Netherlands" if label_country == "The Netherlands"
labmask code_age, val(label_age)
```

```
labmask code_type, val(label_type)
labmask code_country, val(label_country)
labmask code_sex, val(label_sex)
* All countries need to have a 10-year data range...
bys code_type code_country: egen year_max = max(year)
gen year_min = year_max - 9
order year_min, before(year_max)
keep if inrange(year, year_min, year_max)
* Loop for calculating EAPCs by country, age group, and type
gen log_asr = ln(asr_world)
gen aapc = .
gen aapc_lower = .
gen aapc_upper = .
* Generate group variable for the loop...
gsort code_country code_type code_age year
egen group_loop = group(code_country code_type code_age)
levelsof group_loop, local(levels)
foreach i of local levels {
glm log_asr year if group_loop == `i', eform
mat mytable = r(table)
qui replace aapc = mytable[1,1] if group_loop == `i' & missing(aapc)
qui replace aapc_lower = mytable[5,1] if group_loop == `i' & missing(aapc_lower)
qui replace aapc_upper = mytable[6,1] if group_loop == `i' & missing(aapc_upper)
}
replace aapc = (aapc - 1) * 100
replace aapc_lower = (aapc_lower - 1) * 100
replace aapc_upper = (aapc_upper - 1) * 100
drop year asr_world crude_rate cum_risk cases log_asr group_loop
duplicates drop *, force
save "${data}/Manuscript EAPC Data.dta", replace

* Combine the datasets...
use "${data}/Manuscript ASR Data (1998-2017).dta", clear
merge m:1 code_age code_type code_country using "${data}/Manuscript EAPC Data.dta"
drop _merge
merge m:1 label_country using "${data}/Manuscript HDI Data.dta"
drop if _merge != 3
drop _merge
gsort code_country code_type code_age year
save "${data}/Manuscript Full EAPC Data.dta", replace
}

**# UNCERTAINTY LIMITS
/*
*******************************************************************************
TABLE - UNCERTAINTY LIMITS FOR UN/HDIBANK REGIONS
- table of uncertainty limits for ASRs (all ages, under-50, over-50)
*******************************************************************************
*/
{
* Import the dataset for the cases tables.
clear all
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
*** Need to derive 3 sets of ASRs: one for true cases, one for lower UI estimates, one
for upper UI estimates...
* Generate product term for each age group.
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
```

```
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
* Generate estimate-specific crude rates...
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
* Generate the ASRs for each population...
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_region label_region code_type label_type)
gen asr_true = w_true / 100000
gen asr_lower = w_lower / 100000
gen asr_upper = w_upper / 100000

*** Repeat for 0-49 and 50+ ASRs...
* Ages 0-49
clear all
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
keep if code_age <= 10
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_region label_region code_type label_type)
gen asr_true = w_true / 80000
gen asr_lower = w_lower / 80000
gen asr_upper = w_upper / 80000

* Ages 50+
clear all
```

```
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
keep if code_age > 10
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_region label_region code_type label_type)
gen asr_true = w_true / 20000
gen asr_lower = w_lower / 20000
gen asr_upper = w_upper / 20000

*** Need Global ASRs...
** Import the dataset for the cases tables.
clear all
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
drop if code_region > 980
collapse (sum) final_events final_upper final_lower population, ///
by(code_age label_age code_type label_type)
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
```

```
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_type label_type)
gen asr_true = w_true / 100000
gen asr_lower = w_lower / 100000
gen asr_upper = w_upper / 100000

* Ages 0-49...
clear all
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
drop if code_region > 980
keep if code_age <= 10
collapse (sum) final_events final_upper final_lower population, ///
by(code_age label_age code_type label_type)
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_type label_type)
gen asr_true = w_true / 80000
gen asr_lower = w_lower / 80000
gen asr_upper = w_upper / 80000

* Ages 50+
clear all
import excel "${main}/2024 10 31 - Uncertainty Limits for Cases and Deaths by Age
Group.xlsx", ///
case(lower) first sheet(AllAgeGroups)
drop if code_region > 980
keep if code_age > 10
collapse (sum) final_events final_upper final_lower population, ///
by(code_age label_age code_type label_type)
recode code_age (1=0) (2=5) (3=10) (4=15) (5=20) (6=25) (7=30) (8=35) (9=40) (10=45)
(11=50) (12=55) (13=60) (14=65) (15=70) (16=75) (17=80) (18=85)
rename code_age agegr
gen product = .
```

```
replace product = 12000 if agegr == 0
replace product = 10000 if agegr == 5
replace product = 9000 if agegr == 10
replace product = 9000 if agegr == 15
replace product = 8000 if agegr == 20
replace product = 8000 if agegr == 25
replace product = 6000 if agegr == 30
replace product = 6000 if agegr == 35
replace product = 6000 if agegr == 40
replace product = 6000 if agegr == 45
replace product = 5000 if agegr == 50
replace product = 4000 if agegr == 55
replace product = 4000 if agegr == 60
replace product = 3000 if agegr == 65
replace product = 2000 if agegr == 70
replace product = 1000 if agegr == 75
replace product = 500 if agegr == 80
replace product = 500 if agegr == 85
gen crate_true = (final_events / population) * 100000
gen crate_lower = (final_lower / population) * 100000
gen crate_upper = (final_upper / population) * 100000
gen w_true = (crate_true * product)
gen w_lower = (crate_lower * product)
gen w_upper = (crate_upper * product)
collapse (sum) final_events final_upper final_lower (sum) population (sum) w_true w_lower
w_upper, ///
by(code_type label_type)
gen asr_true = w_true / 20000
gen asr_lower = w_lower / 20000
gen asr_upper = w_upper / 20000
}
```